

EduPlanner: LLM-Based Multiagent Systems for Customized and Intelligent Instructional Design

Xueqiao Zhang , Chao Zhang, Jianwen Sun , Jun Xiao , Yi Yang , Senior Member, IEEE, and Yawei Luo 

Abstract—Large language models (LLMs) have significantly advanced smart education in the artificial general intelligence era. A promising application lies in the automatic generalization of instructional design for curriculum and learning activities, focusing on two key aspects: 1) *customized generation*: generating niche-targeted teaching content based on students' varying learning abilities and states and 2) *intelligent optimization*: iteratively optimizing content based on feedback from learning effectiveness or test scores. Currently, a single large LLM cannot effectively manage the entire process, posing a challenge for designing intelligent teaching plans. To address these issues, we developed EduPlanner, an LLM-based multiagent system comprising an evaluator agent, an optimizer agent, and a question analyst, working in adversarial collaboration to generate customized and intelligent instructional design for curriculum and learning activities. Taking mathematics lessons as our example, EduPlanner employs a novel Skill-Tree structure to accurately model the background mathematics knowledge of student groups, personalizing instructional design for curriculum and learning activities according to students' knowledge levels and learning abilities. In addition, we introduce the CIDD_P, an LLM-based 5-D evaluation module encompassing Clarity, Integrity, Depth, Practicality, and Pertinence, to comprehensively assess mathematics lesson plan quality and bootstrap intelligent optimization. Experiments conducted on the GSM8K and Algebra datasets demonstrate that EduPlanner excels in evaluating and optimizing instructional design for curriculum and learning activities. Ablation studies further validate the significance and effectiveness of each component within the framework.

Index Terms—Instructional design, intelligent agent, large language models (LLMs), multiple agents, smart education.

I. INTRODUCTION

IN The educational realm, formulating pedagogical tasks presents a challenge that involves numerous scenarios and

Received 25 July 2024; revised 14 October 2024, 6 January 2025, 17 March 2025, and 2 April 2025; accepted 4 April 2025. Date of publication 16 April 2025; date of current version 28 April 2025. This work was supported in part by the National Natural Science Foundation of China under Grant 62293554 and Grant U2336212, in part by the National Key R&D Program of China under Grant SQ2023AAA01005, in part by the Zhejiang Provincial Natural Science Foundation of China under Grant LZ24F020002, in part by the Ningbo Innovation “Yongjiang 2035” Key Research and Development Program under Grant 2024Z292, and in part by the Young Elite Scientists Sponsorship Program by the China Association for Science and Technology under Grant 2023QNRC001. (Xueqiao Zhang and Chao Zhang contributed equally to this work.) (Corresponding author: Yawei Luo.)

Xueqiao Zhang, Chao Zhang, and Yawei Luo are with the School of Software Technology, Zhejiang University, Ningbo 315048, China (e-mail: yawei@zju.edu.cn).

Jianwen Sun is with the Faculty of Artificial Intelligence in Education, Central China Normal University, Wuhan 430079, China.

Jun Xiao and Yi Yang are with the College of Computer Science and Technology, Zhejiang University, Hangzhou 310007, China.

Digital Object Identifier 10.1109/TLT.2025.3561332

Key Points in Instructional Design for Curriculum and Learning Activities

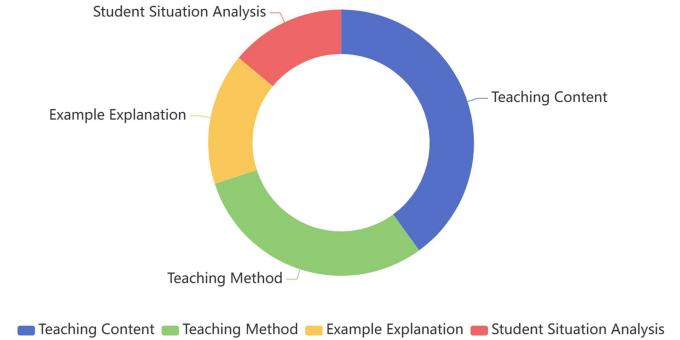


Fig. 1. Proportion of people who consider various factors critical in the task of generating lesson plans. Taking mathematics lessons as our example, in EduPlanner, we model the student situation using a newly proposed Skill-Tree ability module. We iteratively evaluate and optimize the teaching method and teaching content through the adversarial collaboration between the evaluator agent and the optimizer agent. In addition, an analyst agent is employed to analyze error-prone examples and incorporate them into the subsequent lesson plan. EduPlanner effectively addresses the most pressing challenges in lesson planning.

test cases. The crafting of instructional design for curriculum and learning activities [1], [2] is an essential but difficult component of these tasks. It requires a considerable investment of time and energy from educators to understand the diverse learning profiles of their students before making an instructional design for curriculum and learning activities, ensuring a comprehensive structure, clarity of content, the inclusion of appropriate introductions and illustrative examples, as well as reflecting on and optimizing teaching behavior after class. To explore the pedagogical priorities that shape teachers’ instructional designs, Wu [3] revealed that 40.2% educators consider teaching content the paramount aspect, 29.4% place heightened emphasis on the application of teaching methods, and 16.3% focus on example problem explanations. The remaining 14.1% focus on an in-depth analysis of the student situation, as depicted in Fig. 1. Traditionally, teachers have relied on their accumulated experience and innate intuition to devise instructional frameworks. This process often necessitates a trial-and-error approach, with iterative improvements made through classroom implementation. Moreover, the diversity of student dispositions and the unpredictability of the classroom environment compound the difficulty of crafting comprehensive instructional design for curriculum and learning activities prior to instruction. The absence

Question: A brownie recipe is asking for 350 grams of sugar, and a pound cake recipe requires 270 more grams of sugar than a brownie recipe. How many grams of sugar are needed for the pound cake? Final Answer: 620 Section: Translation
Question: Sheila and Lenore were driving to their grandmother's house. Lenore left one hour after Sheila. Sheila drove at a rate of 45 mph, and Lenore drove at a rate of 60 mph. How many hours will it take for Lenore to catchup to Sheila? Final Answer: 3 Section: Solve Applications with Systems of Equations
Question: The perimeter of a rectangle is 84. The length is 10 more than three times the width. Find the length. Final Answer: 34 Section: Solving Systems of Equations by Substitution

Question: Janet 2019s ducks lay 16 eggs per day. She eats three for breakfast every morning and bakes muffins for her friends every day with four. She sells the remainder at the farmers' market daily for \$2 per fresh duck egg. How much in dollars does she make every day at the farmers' market? Solution: Janet sells $16 - 3 - 4 = 9$ duck eggs a day. She makes $9 * 2 = 18$ every day at the farmers' market. Answer: 18
Question: Cynthia eats one serving of icecream every night. She buys cartons of icecream with 15 servings of icecream per carton at a cost of \$4.00 per carton. After 60 days, how much will she spend on icecream? Solution: Each container of icecream has 15 servings and she eats 1 serving a night so after 60 days she will need $60/15 = 4$ containers of icecream. If each carton costs \$4.00 and she needs 4 containers then it will cost her $4*4 = 16.00$. Answer: 16

Fig. 2. Examples of datasets. On the left is GSM8K [17] and on the right is Algebra [18]. GSM8K is a dataset consisting of 8.5K high-quality language-diversified elementary school mathematics word problems released by OpenAI. Algebra is a dataset comprised of 222 examples related to algebraic equations.

of a standard metric to evaluate the quality of these plans further exacerbates the complexity of this task, rendering the entire procedure laborious and often inefficient.

The emerging trend in generative artificial intelligence (GAI), distinguished by its ability to generate a diverse range of content, including texts, audio, images, and videos, has gradually penetrated the realm of education [4], [5], [6], [7], [8], [9], thus influencing various aspects of traditional pedagogy. Recent advances in GAI, such as large language model (LLM), have positioned them as promising solutions for instructional design [10], [11]. LLMs have the ability to develop sophisticated virtual environments and analyze complex social dynamics [12], [13], [14]. By observing agent responses and metric scores, researchers can gain insight into the effectiveness of various strategic approaches within instructional design. However, existing zero-shot instructional design generation methods [15], which depend entirely on the capabilities of LLM, often simply follow instructions to perform tasks without addressing the individual needs of students. In addition, the existing methods [16] only concentrate on the generation and evaluation of instructional designs, yet they fail to incorporate the application of LLM in the optimization process of instructional designs. Consequently, these instructional designs tend to be dull in knowledge transmission, deficient in case explanations, and devoid of a setting in course difficulty levels.

To address these challenges, this article introduces “EduPlanner,” an automated multiagent system designed for the generation, evaluation, and optimization of instructional design. Taking mathematics lessons as our example, “EduPlanner” enables teachers to employ a variety of mathematical instructional design drafts, simulate classroom scenarios, and make targeted adjustments to instructional strategies based on feedback. The system comprises an evaluator agent, an optimizer agent, and a question analyst, which collaborate adversarially to generate customized and intelligent instructional design for curriculum and learning activities.

“EduPlanner” begins by assessing the overall learning situation of the class using an innovative Skill-Tree structure, which informs the creation of teaching resources that align with

individual learning trajectories, thus improving instructional efficacy and student engagement. Based on personal evaluation of the class, “EduPlanner” generates an initial instructional design for curriculum and learning activities and forwards it to the evaluator agent. The evaluator agent employs a 5-D evaluation system comprising Clarity, Integrity, Depth, Practicality, and Pertinence (CIDPP). Feedback from the evaluator is then used by the optimizer agent to improve the instructional design and by the analyst agent to record typical examples. This adversarial cycle mimics a real teacher’s iterative process of instructional design for curriculum and learning activities.

We evaluate the effectiveness of our framework using CIDPP-based experiments on the GSM8K [17] and Algebra [18] datasets, as illustrated in Fig. 2. Compared to conventional methods, EduPlanner significantly improves the efficiency of instructional design for curriculum and learning activities, reduces pedagogical burdens, and facilitates precise alignment with the diverse educational needs of students.

The main contributions of this study can be summarized as follows.

- 1) We introduce “EduPlanner,” a novel framework for customized and intelligent instructional design including three agents: evaluator agent, optimizer agent, and analyst agent.
- 2) We propose an innovative Skill-Tree structure to craft a personalized learning profile for each student, which provides corresponding input for instructional generators, enabling a focus on tailored student needs when structuring educational content.
- 3) We design a comprehensive 5-D instructional design evaluation system (CIDPP), and our framework achieves state-of-the-art results in experiments utilizing this system.

II. RELATED WORK

A. Smart Education

Smart education, a modern campus feature, employs artificial intelligence (AI) and technology to create an interactive learning environment and improve academic achievement [19].

Gao [20] employs AI for speech recognition and customized learning. Almogren et al. [21] use ChatGPT to provide instant assistance, encourage independent learning, and provide prompt feedback. Shrestha et al. [22] present the creation process of the social presence-enabled augmented reality (AR) tool, an innovative AR learning application designed specifically for online engineering education. Pope et al. [23] address the limit of educational tools to teach novice K-12 learners about data-driven systems by introducing the “GenAI Teachable Machine,” a visual data-driven design platform. Babu et al. [24] explore the transformative potential of AI-driven instructional design to enhance personalized learning experiences and improve educational efficiency.

B. GAI for Educational Transformation

GAI has attracted great attention with the advancement of LLMs. Within the educational domain, GPTEach [25] employs GPT to simulate student behaviors. EduAgent [10] excels in anticipating the detailed learning actions of students and generating plausible learning behaviors for virtual students without ground truth. AgentVerse [12] enables the development, training, and interaction of agents across various domains, simulating complex human decision-making processes. Gradescope [26] improves the assessment process for educators by organizing assessment tasks, increasing student engagement, and providing valuable data on student performance.

C. LLM-Based Multiagent Systems

Recent studies have explored the use of LLMs in agent systems, demonstrating excellent performance in accomplishing tasks [14], [27] upon prompts through collaborative interactions among multiple agents. The concept of generative agents [13] involves the creation of a simulated town with 25 agents. The approach relies on multiple language models [28], [29] and often proves inefficient, particularly in open-ended settings. Managing the control aspects of LLM agents, Zheng et al. [30] present a challenge, yet this is where reinforcement learning [31], [32] excels. MetaGPT [33] uses standard operating procedures to enhance the multiagent systems and emulates a software team with agents including product manager, architect, project manager, engineer, and quality engineer. Autogen [34] offers designable conversable agents and is capable of integrating with external tools to facilitate collaborative conversations within multiagent systems. AutoGPT [35] works based on established goals without human intervention and employs a real-time loop evaluation strategy to assess the completion of objectives dynamically. CAMEL [36] develops a role-playing framework that generates dialogue datasets to foster communication and collaboration within multiagent systems.

D. Evaluation Using LLMs

Most of the evaluations in relative experiments are open-ended and lack definitive reference answers; designing a rule-based program to evaluate the output of LLM is challenging. Traditional evaluation metrics based on the similarity between

the output and the reference answers (e.g., ROUGE [37] and BLEU [38]) also fail to address these questions. Employing experts for manual evaluation is time-consuming and expensive. Therefore, in order to facilitate an efficient automated evaluation process, some works explore LLMs as a substitute for human evaluation. These models, often trained via reinforcement learning from human feedback, have demonstrated great alignment with humans. The LMSys [39] explores the merits and limitations of utilizing diverse LLMs as judges for tasks ranging from writing and mathematics to general world knowledge. The self-rewarding LLM [40] trains itself through the data generated by its own and evaluated by the GPT-4. CharacterLLM [41] employs GPT as an evaluator to access the LLMs in role-playing scenarios, and Neeko [42] evaluates the performance on role switching of LLM in multirole-playing contexts based on GPT. L-eval [43] takes the pairwise battle format and reports the win rate versus LLMs, augmenting the judgment prompt to mitigate the preference for more detailed and lengthy answers.

III. METHODOLOGY

A. Overview

We present EduPlanner, a multiagent-based system designed for the evaluation and optimization of instructional designs. The system employs a modular architecture, depicted in Fig. 3. To tailor instructional design for specific student groups, we model students’ knowledge backgrounds using a Skill-Tree structure. This structure facilitates the generation and optimization of instructional designs that meet the diverse learning needs of students. Our approach includes three collaborative agents: the evaluator agent, which assesses the quality of instructional designs; the optimizer agent, which improves lesson content; and the analyst agent, which identifies common error points in example questions within the instructional designs. These agents are integrated using an adversarial learning framework. The evaluator agent provides feedback on the advantages and disadvantages of the instructional designs, guiding the optimizer agent and analyst agent. The evaluation scores serve as targets for optimization, driving an iterative adversarial process that enhances the instructional design for curriculum and learning activities. When optimization is complete, the instructional design with the highest score in the queue is output and its content mainly includes two parts: knowledge point explanation and question explanation.

B. Skill-Tree Structure

To generate customized instructional design for students with different knowledge backgrounds, a key issue to be addressed is the modeling of students’ abilities. In configurable general multi-agent interaction (CGMI) [11], a character tree structure was designed to solve the problem of inconsistency of roles in LLM after multiple rounds of dialogue . Following CGMI [11], we choose to employ a tree structure to portray the competencies¹ of students. In a novel extension of this approach, we

¹We use mathematics as an example to illustrate the role of the Skill-Tree structure, while other subjects can also be applied.

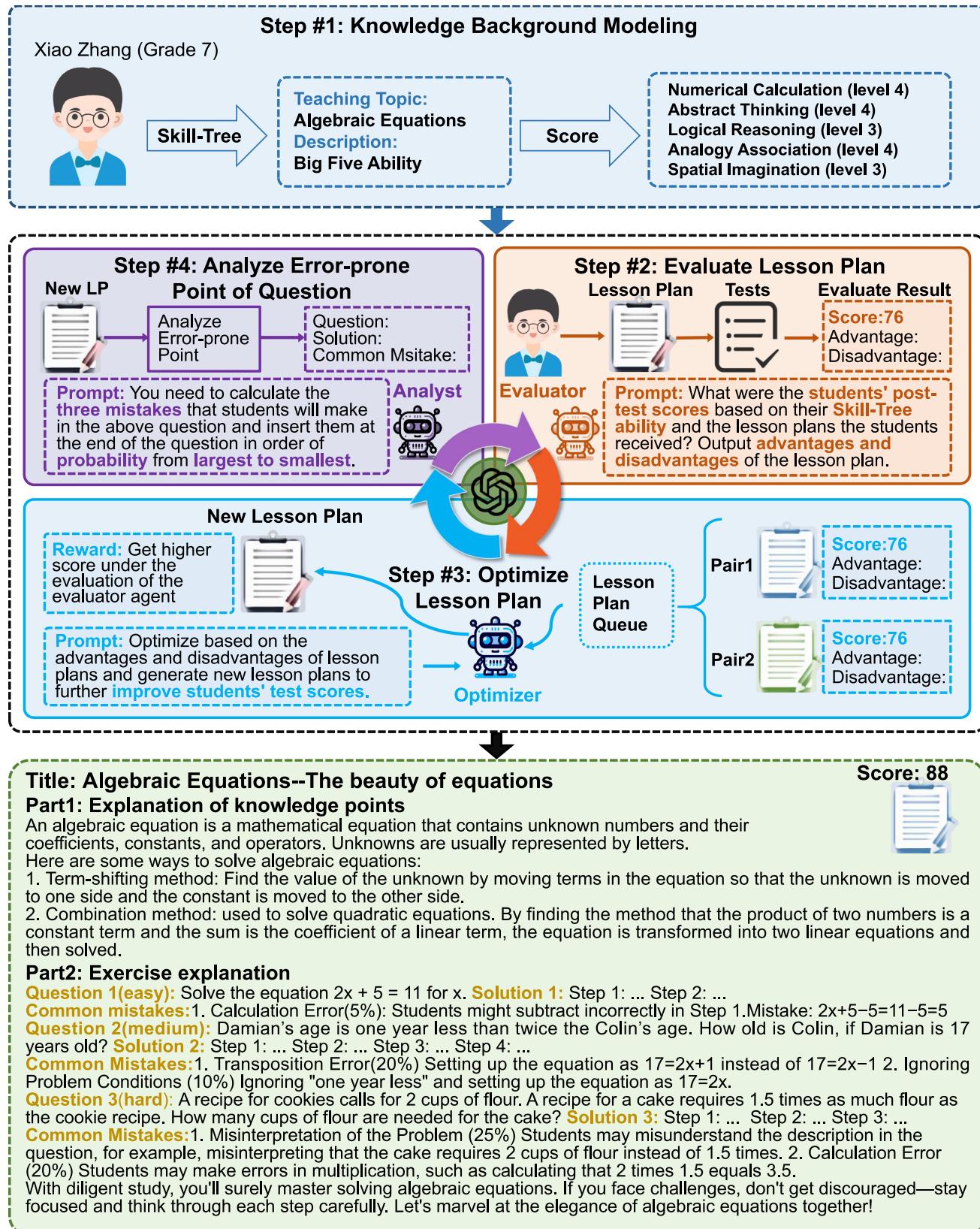


Fig. 3. EduPlanner: LLM multiagent-based instructional design generation framework. When the optimization is completed, the instructional design for curriculum and learning activities with the highest score in the queue is output, and its content mainly includes two parts: knowledge point explanation and question explanation.

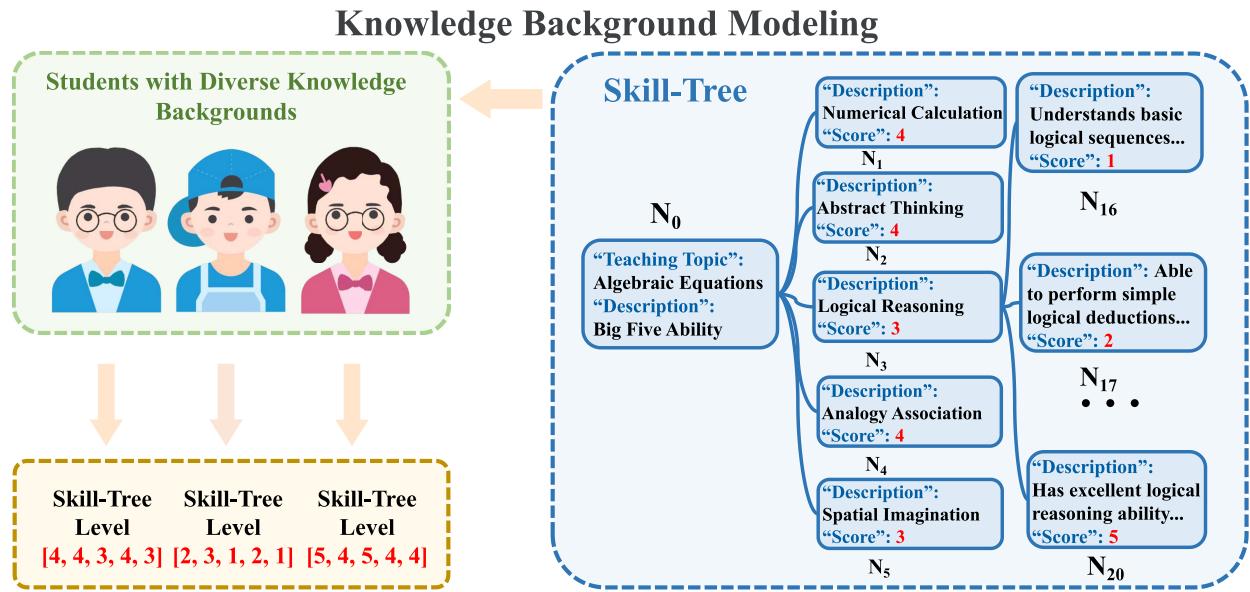


Fig. 4. Skill-Tree structure models the students' diverse knowledge background through five principal abilities, outputs the students' Skill-Tree level, and provides it to the agents so that the agents can capture the knowledge background of different students and ensure that targeted instructional design for curriculum and learning activities can be generated.

introduce five specialized subnodes under each primary node. This enhancement enables a detailed quantification of abilities across different dimensions, offering an in-depth description of the knowledge bases and learning abilities of individual students, as shown in Fig. 4. Compared to traditional binary evaluations (good or bad), this approach offers a more comprehensive analysis with multiple dimensions, enabling the evaluator agent to capture the differences in students' abilities.

Student abilities (mathematics as example):

- 1) *Node N₀*: It explains the content of the teaching topic and the description of nodes N₁–N₅
- 2) *Node N₁*: It represents Numerical Calculation ability, reflecting the accuracy and proficiency of students in addition, subtraction, multiplication, and division.
- 3) *Node N₂*: It represents Abstract Thinking ability, one of the key factors in understanding and solving algebraic equations.
- 4) *Node N₃*: It depicts Logical Reasoning ability, a core skill for solving mathematical problems.
- 5) *Node N₄*: It represents Analogy Association ability, an important tool for students to expand knowledge boundaries.
- 6) *Node N₅*: It refers to Spatial Imagination ability, which is particularly important for solving geometric problems in algebraic equations.

C. Instructional Design Evaluation

In educational practice, the effectiveness of instructional design is a challenging issue. Given the great performance on various tasks of LLM-Based-Agents, we are encouraged to utilize agents as experienced evaluation experts that provide feedback on instructional designs, such as advantages, disadvantages, and suggestions for improvement.

1) **Evaluation Criteria:** To comprehensively evaluate instructional designs and catalyze intelligent optimization, we introduce the CIDDP, an LLM-based 5-D evaluation module encompassing Clarity, Integrity, Depth, Practicality, and Pertinence. This evaluation system comprises the pedagogical philosophy in instructional design, with each dimension serving a distinct purpose.

- 1) **Clarity** is essential for making instructional designs clear and removing unnecessary information. It helps teachers establish clear teaching goals, ensuring that educational activities are aligned with pedagogical principles of purposeful instruction.
- 2) **Integrity** involves examining instructional designs to ensure that they are detailed and complete, covering all necessary aspects like clear explanations of concepts and relevant examples. It aligns with the pedagogical philosophy that promotes a comprehensive and integrated approach to learning.
- 3) **Depth** measures how instructional designs engage students in deep thinking and present content that reveals the deep meanings and relationships within knowledge. It is in line with the pedagogical emphasis on fostering deep learning experiences.
- 4) **Practicality** assesses the real-world applicability of instructional designs and how well students can use their knowledge to solve practical problems. It corresponds with the pedagogical philosophy that focuses on applying learned knowledge.
- 5) **Pertinence** examines the efficacy of instructional designs in satisfying the diverse learning needs of students. It promotes the adaptation of different levels of knowledge and learning preferences, reflecting the personalized teaching approach central to modern pedagogical thought.

Algorithm 1: Evaluator Agent Expert Evaluation.

Input: Evaluator Agent A_e , Initial Instruction Design \mathcal{L}_{p_0} , Knowledge Background, Evaluation Task, Test Questions $\mathcal{X}_1, \mathcal{X}_2, \dots, \mathcal{X}_T$

Output: post score s , advantage a , disadvantage d

- 1 Set knowledge background || Instruction Design || $\mathcal{X}_1 \rightarrow \mathcal{P}_1$
- 2 Give \mathcal{P}_1 as input to A_e and receive output \mathcal{O}_1
- 3 **for** i in T **do**
- 4 Set \mathcal{X}_i || evaluation task $\rightarrow \mathcal{P}_i$
- 5 Give \mathcal{P}_i as input to A_e and receive output \mathcal{O}_i
- 6 Extract s_1, \dots, s_T from $\mathcal{O}_1, \dots, \mathcal{O}_T$
- 7 Set $\sum_i^T s_i / T \rightarrow s$
- 8 Extract $a_1, \dots, a_T, d_1, \dots, d_T$ from $\mathcal{O}_1, \dots, \mathcal{O}_T$
- 9 Use A_e to summarize $a_1, \dots, a_T \rightarrow a, d_1, \dots, d_T \rightarrow d$
- 10 return (s, a, d)

2) *Evaluator Agent:* The evaluator agent, built on the Meta-Llama-3-70B LLM [44], is designed to evaluate instructional designs and offer feedback. Trained on 100 carefully annotated instructional design for curriculum and learning activities, each assessed by education experts for effectiveness, scoring, advantages, and disadvantages, the evaluator agent demonstrates an evaluation capability comparable to that of human experts. The process involves analyzing students' Skill-Tree scores to measure their knowledge levels and accessing scores after instruction of instructional designs using randomly selected test questions. The agent provides detailed feedback that clearly points out the advantages and content that needs improvement, forming a crucial basis for the subsequent optimization of the instructional design.

3) *Instructional Design Evaluation Algorithm:* The evaluator agent assesses the quality of instructional design and provides feedback, denoted as F (e.g., advantages and disadvantages), to the optimizer agent. We refer to this algorithm as evaluator agent expert evaluation (EAEE) [15]. Specifically, we allow the evaluator agent understand students' background knowledge and learning abilities based on their Skill-Tree scores. Then, we provide the students with a specific instructional design for curriculum and learning activities and let the evaluator agent evaluate the students' average scores on exam questions. The specific mathematical equation is as follows:

$$R(lp, S) = \frac{1}{|T|} \sum_{i=1}^{|T|} f(lp, S, x_i) \quad (1)$$

where x_i represents the i th test question, lp is the instructional design to be evaluated, and S is the Skill-Tree score of the student, allowing the LLMs to capture differences in students' learning abilities. The function $f(\cdot)$ is the evaluation function of the evaluator agent. T is a hyperparameter we set, which represents the number of test questions used during an instructional design for evaluation. $R(\cdot)$ represents the average test score evaluated by the evaluator agent for students with test questions. The evaluator agent provides feedback on the instructional design for curriculum and learning activities. Algorithm 1 details the EAEE process.

The process of the EAEE algorithm is as follows. The student role description, instructional design, evaluation task description, and the first test question are combined to form the initial prompt input p_1 . The initial input p_1 is submitted to the evaluator agent, and its output \mathcal{O}_1 is received, which includes predictions of the student's learning effectiveness. The process is repeated for the M test questions. Finally, the model extracts the predicted probabilities or scores for the student's learning success from each output and calculates the average. The advantages and disadvantages of the instructional design generated in each output are also extracted and summarized by A_e . The summary is provided to the optimizer agent A_o as reference information for optimization of the instructional design.

D. Instructional Design Optimization

1) *Optimizer Agent:* The optimizer agent is designed to optimize personalized high-quality instructional designs according to the students' knowledge backgrounds. The process involves using the feedback F provided by the evaluator agent (i.e., the advantages and disadvantages of the instructional design) for optimization. The goal is to maximize the instructional design's evaluation score assessed by the evaluator agent. This instructional design optimization strategy is expressed as follows:

$$\begin{cases} F = f(lp, S) \\ lp_{\text{new}} = \arg \max_{R(lp_{\text{new}}, S)} \{g(D, F)\}. \end{cases} \quad (2)$$

In this equation, F represents the feedback generated by the evaluator agent, the function $g(\cdot)$ is responsible for optimizing and generating a new instructional design using the feedback and the instructional designs in the queue D along with their scores, and $R(\cdot)$ represents the evaluation score for the new instructional design. The final generated lp_{new} is optimized to maximize the instructional design's evaluation score.

Our optimizer agent employs a dynamically sorted instructional design queue to prioritize the most effective instructional design. After the evaluator agent evaluates the new instructional design, which involves generating scores and feedback, the optimizer agent inserts the instructional design for the curriculum and learning activities into the queue based on its score. The process of the instructional design queue is described as follows:

$$D_n = \text{Trim} (\text{Sort} (\text{Ins} (lp_n, D), R(lp_n, S)), P) \quad (3)$$

where D denotes the instructional design queue, lp_n is the new instructional design, S represents the Skill-Tree scores of the students, P indicates the maximum capacity of the instructional design queue, $R(\cdot)$ represents the student's average test score, $\text{Ins}(\cdot)$ indicates inserting the new instructional design into the instructional design queue, $\text{Sort}(\cdot)$ means sorting the queue in ascending order, and $\text{Trim}(\cdot)$ keeps only the top P instructional design.

2) *Instructional Design Optimization Algorithm:* We refer to the instructional design optimization algorithm as optimizer agent expert optimization [15]. The agent optimizer (A_o) and the agent example analyzer (A_a) form the core of the optimization process. The student description, initial instructional design, evaluation task, optimization task, and the first test question are

Algorithm 2: Optimizer Agent Expert Optimization.

Input: Evaluator Agent \mathcal{A}_e , Optimizer Agent \mathcal{A}_o , Question Analyst \mathcal{A}_a , Initial Instructional Design $\mathcal{L}p_0$, Knowledge Background \mathcal{K} , Evaluation Task, Optimization Task, Test Questions $\mathcal{X}_1, \mathcal{X}_2, \dots, \mathcal{X}_T$, Optimization Rounds \mathcal{N}

Output: Instructional Design $\mathcal{L}p_0, \mathcal{L}p_1, \dots, \mathcal{L}p_{p-1}$,

- 1 Evaluate $\mathcal{L}p_0$ using Algorithm 1 and receive score \mathcal{R}_0
- 2 Set $(\mathcal{L}p_0, \mathcal{R}_0) \rightarrow \mathcal{D}$
- 3 Set knowledge background optimization task $\rightarrow \mathcal{P}_0$
- 4 **for** i in N **do**
- 5 Set $[] \rightarrow D'$
- 6 **for** k in \mathcal{K} **do**
- 7 Generate $\mathcal{C}_{n-1}\mathcal{K}+k$ by feeding \mathcal{P}_{n-1} into \mathcal{A}_o
- 8 Insert q into $\mathcal{C}_{n-1}\mathcal{K}+k$
- 9 Evaluate $\mathcal{C}_{n-1}\mathcal{K}+k$ using Algorithm 1 and receive score $\mathcal{R}_{n-1}\mathcal{K}+k$
- 10 Analysis error-prone points of questions in $\mathcal{C}_{n-1}\mathcal{K}+k$ using Algorithm 3 and receive new $\mathcal{C}_{n-1}\mathcal{K}+k$
- 11 Set $D' \cup (\mathcal{C}_{n-1}\mathcal{K}+k, \mathcal{R}_{n-1}\mathcal{K}+k) \rightarrow \mathcal{D}$
- 12 D.append(D')
- 13 Sort D by scores in ascending order
- 14 Set $D[: \text{max length}] \rightarrow D$
- 15 Set knowledge background || D || optimization task $\rightarrow \mathcal{P}_n$
- 16 **return** \mathcal{D}

combined into the initial prompt (p_0). This prompt is submitted to \mathcal{A}_o , which generates an optimized instructional design for curriculum and learning activities output (o_0), including specific suggestions for content improvement.

The process is repeated for each test question. \mathcal{A}_a extracts examples from each output, forming a set of examples. All suggestions and examples are combined into the final instructional design for curriculum and learning activities. Algorithm 2 provides a detailed description of the optimization process.

E. Instructional Design Analysis

1) *Analyst Agent*: The analyst agent explains the common mistakes produced by questions in the instructional design to alert students. We compile 50 common mistakes in algebraic equations into five groups. For each example from Algorithm 2, the analyst agent identifies three common error-prone points for students with corresponding knowledge backgrounds. These error points are inserted into the example explanations to help students avoid these errors.

2) *Error-Prone Point Analysis Algorithm*: To identify common errors students may make, we develop the analyst agent expert analysis algorithm. The algorithm extracts examples from instructional designs. Utilizing annotated common errors as reference, the analyst agent generates potential errors that students might make for each example. These errors are inserted into the example explanations within the instructional design, serving as reminders to help students avoid these mistakes and correctly solve the problems. Algorithm 3 provides a detailed description of this process.

Algorithm 3: Analyst Agent Expert Analysis.

Input: Question Analyzer \mathcal{A}_a , Common Mistakes Database (CMD), Instructional Design (Lp), Example Questions (EQs) generated by Optimizer Agent, Knowledge Background (KB)

Output: new instructional designs with integrated error-prone points and probabilities(nlp)

- 1 $A_a \leftarrow$ Initialize Analyst Agent
- 2 $EQs \leftarrow$ Retrieve example questions from Lp
- 3 Load CMD
- 4 **for** eq in EQs **do**
- 5 $KB \leftarrow$ Identify target student group knowledge background for eq
- 6 $CMs \leftarrow$ Retrieve top 3 common mistakes from CMD based on KB and mistake probability
- 7 **for** cm in CMs **do**
- 8 $mp \leftarrow$ Calculate mistake probability for eq
- 9 $me \leftarrow$ Generate explanation for cm
- 10 $CMs \leftarrow$ Sort CMs by mp in descending order
- 11 Insert CMs and me into eq section of Lp
- 12 **return** nlp

IV. EXPERIMENTS

In this section, we verify the results of the generation framework. We propose an instructional design evaluation system (CIDPP) to analyze the quality of instructional designs. In addition, we conducted ablation experiments to demonstrate the roles of each module in the framework.

A. Hyperparameter Configuration

In this section, we tune the hyperparameters of each component to ensure optimal performance. For the evaluator agent's base model, i.e., Meta-Llama-3-70B-Instruct, we set the temperature parameter at 0.0 for stable output. The optimizer agent's base model, i.e., GPT-4, has a temperature parameter set at 1.0 for diverse instructional design content. The analyzer, also based on GPT-4, has a temperature parameter set at 0.7 for balance between stability and diversity.

B. Contrastive Results

To validate the efficacy of our framework, we compare our results with other methods by LLMs like GPT-3.5-turbo, Llama-3-70B-Instruct, and GPT-4, as detailed in Fig. 5. For the direct generation of instructional designs through LLMs, we apply a prompt as follows: “*Generate an instructional design with the teaching theme of algebraic equations. It contains two parts, the explanation of knowledge points and the explanation of examples. The explanation of knowledge points should be comprehensive, and the explanation of examples should be representative. The generated instructional designs must be concise and limited to 200 words.*” Finally, we assess the instructional design based on the CIDPP, as shown in Table I, where the results objectively demonstrate the effectiveness of our framework in improving the instructional designs. The prompt of evaluation is as follows.

“*Role: You are an impartial evaluator, experienced in educational content analysis and instructional design evaluation.*

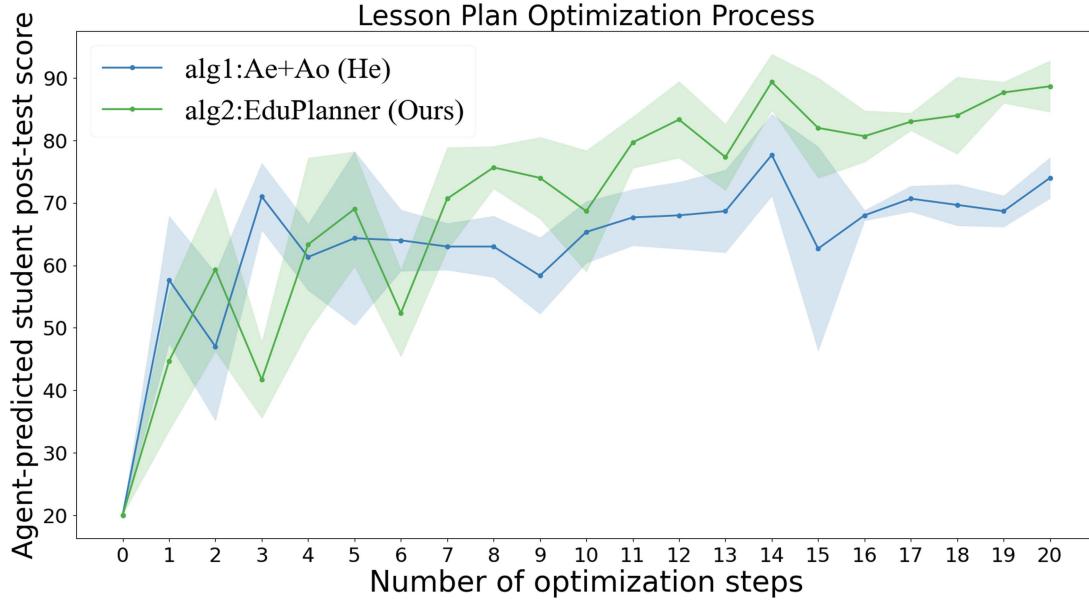


Fig. 5. Optimization process curves of our framework and He-Yueya et al.’s method [15] under the premise that the hyperparameters are consistent with the initial instructional design for curriculum and learning activities. Our framework has a higher score in the instructional design than He-Yueya et al.’s [15] method, and the optimization process is smoother.

TABLE I
QUALITY INDICATORS

Strategy	Clarity↑	Integrity↑	Depth↑	Practicality↑	Pertinence↑	Score↑
GPT-3.5-turbo	28.2	28.1	33.6	33.8	10.0	17
Llama-3-70B-Instruct	35.2	20.7	40.9	30.4	15.8	20
GPT-4	52.9	61.7	42.0	54.4	31.0	49
EduPlanner	84.1	87.5	88.9	87.8	83.7	88

Attention: You are responsible for assessing the quality of a given instructional design based on five specific evaluation criteria. Your evaluation should be objective and based solely on the Evaluation Standard provided below. Lesson Plan: {lessonplan} Evaluation Standard: Clarity: The lesson plan’s directness and simplicity, ensuring it avoids unnecessary complexity and redundancy. Integrity: Whether the lesson plan is complete and systematic, covering both knowledge point explanations and exercise explanations in a complementary manner. Depth: The ability of the lesson plan to inspire deep thinking and facilitate understanding of the underlying connections between knowledge points. Practicality: The practical application value of the examples in the lesson plan, ensuring students can use the knowledge to solve real-life problems. Pertinence: The adaptability of the lesson plan to different students’ knowledge levels and learning needs to achieve optimal learning outcomes. Constraints: Avoid any bias in evaluation based on the content’s length or appearance. Be as objective as possible in assessing each aspect individually without favoring any specific structure or terminology. Workflow: Output your final verdict in the following format: “[A]: [points]; [short analyzes]”, “[B]: [points]; [short analyzes]”, “[C]: [points]; [short analyzes]”, “[D]: [points]; [short analyzes]”, “[E]: [points]; [short analyzes]”. Take a deep breath and think step by step!”

In our research, we introduce a novel framework designed to enhance the optimization of instructional design for curriculum

and learning activities. Based on baseline [15], we preserve the evaluator and optimizer agent, while integrating two crucial additions: a Skill-Tree structure and an analyst agent. As illustrated in Fig. 5, our method outperforms baseline [15] and exhibits a smoother process.

Furthermore, we conduct an ablation study on all components of our instructional design in an optimization framework. As shown in Table II and Figs. 6 and 7, we evaluate the individual impact of each component and measure its contribution to the optimization results using CIDPP.

These contrastive results emphasize that the integration of the evaluator agent, optimizer agent, analyst agent, Skill-Tree mechanisms, and CIDPP within our methodology improves the structural integrity of instructional design and enhances their quality and pedagogical resonance through personalized improvements. Our experiments have shown that utilizing a stand-alone evaluator agent or an optimizer agent does not produce the most favorable results. However, integrating these agents establishes a great baseline, excelling in the dimensions of clarity, integrity, and depth. Moreover, the addition of an analyst agent has notably improved the practicality and pertinence of instructional design for curriculum and learning activities. Although the enhancement in clarity may not be pronounced, the Skill-Tree mechanism improves other performance metrics by creating a map of the student’s competencies and crafting educational content that is aligned with their individual knowledge background.

TABLE II
ABLATION STUDY ON THE GSM8K AND ALGEBRA DATASETS

Strategy	Clarity↑	Integrity↑	Depth↑	Practicality↑	Pertinence↑	Score↑
Ae	35.2	20.7	40.9	30.4	15.8	20
Ao	52.9	61.7	42.0	54.4	31.0	49
Ae+Ao (He [15])	85.3	76.3	78.7	68.5	31.9	74
Ae+Ao+Aa	84.9	75.1	75.9	85.1	82.1	79
Skill-Tree+Ae+Ao	85.4	83.2	78.8	69.8	75.3	84
Skill-Tree+Ae+Ao+Aa (EduPlanner)	84.1	87.5	88.9	87.8	83.7	88

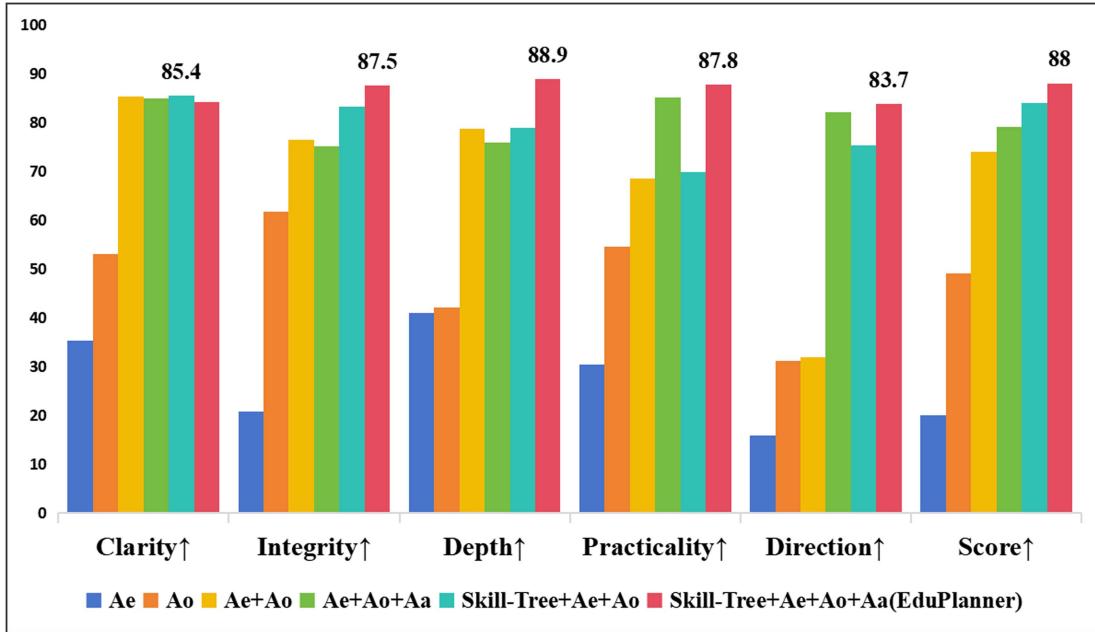


Fig. 6. Bar chart depicting the influence of each framework component on each evaluation metric in CIDDP.

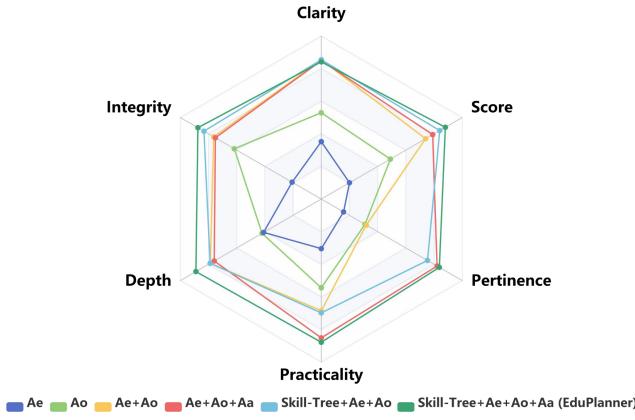


Fig. 7. Radar chart portraying the incremental effects of each component.

V. DISCUSSION

The advent of artificial general intelligence (AGI) has ushered a transformation across diverse sectors, presenting both opportunities and challenges for the field of education. It extends novel forms of support for teaching and learning, introducing innovative methodologies to educational research with a focus

on multidisciplinary integrated approaches like pedagogical [45] and AI. Smart education [46], [47], [48], [49], identified as a pivotal trend in the future educational landscape, is obtaining increasing attention and importance. Central to this trend is personalized education, which disrupts traditional teacher-student roles by dynamically adjusting educational resources and teaching strategies to meet individual students' needs.

Developing instructional design for curriculum and learning activities is a crucial aspect of educational practice that must adapt to changing requirements. The traditional method of creating instructional designs is time consuming for teachers, which involves understanding each student's individual characteristics and knowledge level to effectively align subject matter with their needs. As educational concepts evolve, previous teaching methods may no longer meet the needs of current students. Developing effective instructional design requires the integration of clear assessment methods and feedback mechanisms for quick adjustments and improvements in teaching strategies, which still presents a challenge. Modern instructional design often involves using educational technologies like multimedia tools and online platforms. Educators face the added challenge of mastering these tools and integrating them into their teaching practices.

Considering the existing issues, our framework introduces an evaluative system for instructional designs, a Skill-Tree model for student ability, and a structure with three collaborative agents. These advancements facilitate a comprehensive evaluation and improvement of instructional design quality, customized to address individual student needs.

In our studies, we propose a novel evaluation system for instructional designs that considers clarity, integrity, depth, practicality, and pertinence. This evaluation system reflects the pedagogical principles in instructional design. The instructional design's clarity helps teachers establish clear teaching objectives, ensuring that teaching activities are purposeful and in line with goal-oriented teaching principles in pedagogical philosophy. Ensuring the integrity of instructional designs emphasizes essential elements such as detailed explanations of key concepts and relevant examples, aligning with the comprehensive learning principles promoted in educational philosophy. The depth of instructional designs measures how well the instructional design engages students in deep thinking and provides detailed explanations of the deep interrelations of knowledge, aligning with the pedagogical emphasis on deep-level learning. The practicality of instructional designs evaluates how effectively students can apply their knowledge to solve real-world problems. It corresponds with the pedagogical philosophy that emphasizes the application of knowledge to address practical issues. The pertinence focuses on tailoring instructional designs to meet students' varying knowledge levels and learning requirements, aligning with the personalized teaching philosophy promoted by pedagogical philosophy.

VI. CONCLUSION

In this article, we develop EduPlanner, a multiagent system comprising three key intelligent agents, i.e., the evaluator, the optimizer, and the example analyzer, for the automated evaluation and optimization of instructional designs. To meet students with varying knowledge levels, we design a Skill-Tree structure to measure their knowledge bases, which allows the system to capture the differences in each student's learning abilities and generate corresponding instructional designs. Through a series of experiments, we validate our framework's superior performance over other methods. We also conducted ablation experiments to examine the roles of the Skill-Tree structure and the analyst agent. The results confirm that these novel components are crucial in generating remarkable instructional designs that show the transformative potential of AGI to address complex problems and complex challenges in the domain of education.

VII. LIMITATIONS AND FUTURE WORK

A. Limitations

While EduPlanner has achieved notable results in the generation and improvement of instructional design, it faces various limitations due to the constantly changing educational technology environment and the increasing diversity of teaching needs.

1) The skill tree is specifically tailored to generate instructional designs within certain subjects, which may limit its applicability to other disciplines.

2) EduPlanner's reliance on LLM resources is marked by substantial consumption, which leads to a huge cost in terms of time and money.

3) There is an absence of adaptive integration with existing tools within practical teaching environments, which indicates a demand for ensuring alignment with current educational practices.

B. Future Work

GAI is revolutionizing instructional designs evaluation and optimization, aiming to better address the evolving needs of educational systems [50], [51], [52], [53], [54]. Despite current challenges, continuous improvements in these systems are expected to boost the educational content. Future research should focus on creating models tailored to specific subjects and improving LLMs to better replicate real-world educational settings. Another key strategy involves the combination of various disciplines, improving students' skills in integrating knowledge across different subjects. The knowledge developed through optimizing instructional designs influences other smart education domains. Incorporating LLMs into instructional designs can enhance the adaptability and effectiveness of educational practices in response to the dynamic educational environment.

REFERENCES

- [1] M. Cevikbas, J. Koenig, and M. Rothland, "Empirical research on teacher competence in mathematics lesson planning: Recent developments," *ZDM—Math. Educ.*, vol. 56, no. 1, pp. 101–113, 2024.
- [2] S. Liu, Q. Li, X. Shen, J. Sun, and Z. Yang, "Automated discovery of symbolic laws governing skill acquisition from naturally occurring data," *Nat. Comput. Sci.*, vol. 4, no. 5, pp. 334–345, 2024.
- [3] Y. Wu, "A research on teaching management of teachers in Tianzhi Normal University affiliated technical school," Master's Thesis, School Vocational Educ. Tianjin Univ. Technol., Tianjin, China, 2018.
- [4] C. C. Tossell, N. L. Tenhundfeld, A. Momen, K. Cooley, and E. J. de Visser, "Student perceptions of ChatGPT use in a college essay assignment: Implications for learning, grading, and trust in artificial intelligence," *IEEE Trans. Learn. Technol.*, vol. 17, pp. 1069–1081, 2024.
- [5] S. Liu et al., "COMET: "Cone of experience" enhanced large multimodal model for mathematical problem generation," *Sci. China Inf. Sci.*, vol. 67, no. 12, pp. 1–2, 2024.
- [6] Z. Liu, S. X. Yin, C. Lee, and N. F. Chen, "Scaffolding language learning via multi-modal tutoring systems with pedagogical instructions," in *IEEE Conf. Artif. Intell.*, 2024, pp. 1258–1265.
- [7] S. Ma, Y. Luo, and Y. Yang, "Personas-based student grouping using reinforcement learning and linear programming," *Knowl.-Based Syst.*, vol. 281, 2023, Art. no. 111071.
- [8] J. Sun, F. Yu, S. Liu, Y. Luo, R. Liang, and X. Shen, "Adversarial bootstrapped question representation learning for knowledge tracing," in *Proc. 31st ACM Int. Conf. Multimedia*, 2023, pp. 8016–8025.
- [9] R. Zou et al., "Balancing humans and machines: A study on integration scale and its impact on collaborative performance," in *Proc. AAAI Conf. Artif. Intell.*, 2024, pp. 17628–17636.
- [10] S. Xu, X. Zhang, and L. Qin, "EduAgent: Generative student agents in learning," 2024, *arXiv:2404.07963*.
- [11] S. Jinxin, Z. Jiabao, W. Yilei, W. Xingjiao, L. Jiawen, and H. Liang, "CGMI: Configurable general multi-agent interaction framework," 2023, *arXiv:2308.12503*.
- [12] W. Chen et al., "AgentVerse: Facilitating multi-agent collaboration and exploring emergent behaviors in agents," in *Proc. 12th Int. Conf. Learn.*, 2023.

- [13] J. S. Park, J. O'Brien, C. J. Cai, M. R. Morris, P. Liang, and M. S. Bernstein, "Generative agents: Interactive simulacra of human behavior," in *Proc. 36th Annu. ACM Symp. User Interface Softw. Technol.*, 2023, pp. 1–22.
- [14] C. Qian et al., "Communicative agents for software development," 2023, *arXiv:2307.07924*.
- [15] J. He-Yueya, N. D. Goodman, and E. Brunskill, "Evaluating and optimizing educational content with large language model judgments," in *Proc. 17th Int. Conf. Educ. Data Mining*, 2024, pp. 68–82.
- [16] B. Hu, L. Zheng, J. Zhu, L. Ding, Y. Wang, and X. Gu, "Teaching plan generation and evaluation with GPT-4: Unleashing the potential of LLM in instructional design," *IEEE Trans. Learn. Technol.*, vol. 17, pp. 1445–1459, 2024.
- [17] K. Cobbe et al., "Training verifiers to solve math word problems," 2021, *arXiv:2110.14168*.
- [18] J. He-Yueya, G. Poesia, R. E. Wang, and N. D. Goodman, "Solving math word problems by combining language models with symbolic solvers," in *Proc. 3rd Workshop Math. Reasoning AI at NeurIPS'23*, 2023.
- [19] H. U. Khan, M. Abbas, O. Alruwaili, S. Nazir, M. H. Siddiqi, and S. Alanazi, "Selection of a smart and secure education school system based on the Internet of Things using entropy and TOPSIS approaches," *Comput. Hum. Behav.*, vol. 159, 2024, Art. no. 108346.
- [20] B. Gao, "Understanding smart education continuance intention in a delayed benefit context: An integration of sensory stimuli, UTAUT, and flow theory," *Acta Psychol.*, vol. 234, 2023, Art. no. 103856.
- [21] A. S. Almogren, W. M. Al-Rahmi, and N. A. Dahri, "Exploring factors influencing the acceptance of ChatGPT in higher education: A smart education perspective," *Heliyon*, vol. 10, 2024, Art. no. e31887.
- [22] S. Shrestha, Y. Shan, R. Emerson, and Z. Hosseini, "Developing and usability testing of an augmented reality tool for online engineering education," *IEEE Trans. Learn. Technol.*, vol. 18, pp. 13–24, 2024.
- [23] N. Pope, J. Kahila, H. Vartiainen, and M. Tedre, "Children's AI design platform for making and deploying ML-driven apps: Design, testing, and development," *IEEE Trans. Learn. Technol.*, vol. 18, pp. 130–144, 2025.
- [24] C. S. Babu, M. Bhavesh, A. Gururaja, and T. Venkatesh, "AI-driven instructional design: Enhancing learning experiences through personalization and efficiency," in *Effective Instructional Design Informed by AI*. Hershey, PA, USA: IGI Global, 2025, pp. 87–114.
- [25] J. M. Markel, S. G. Opferman, J. A. Landay, and C. Piech, "GPTeach: Interactive TA training with GPT-based students," in *Proc. 10th ACM Conf. Learn. Scale*, 2023, pp. 226–236.
- [26] A. Singh, S. Karayev, K. Gutowski, and P. Abbeel, "Gradescope: A fast, flexible, and fair system for scalable assessment of handwritten work," in *Proc. 4th ACM Conf. Learn. Scale*, 2017, pp. 81–88.
- [27] J. Li et al., "Agent hospital: A simulacrum of hospital with evolvable medical agents," 2024, *arXiv:2405.02957*.
- [28] G. Wang et al., "Voyager: An open-ended embodied agent with large language models," *Trans. Mach. Learn. Res.*, 2023.
- [29] Y. Luo and Y. Yang, "Large language model and domain-specific model collaboration for smart education," *Front. Inf. Technol. Electron. Eng.*, vol. 25, no. 3, pp. 333–341, 2024.
- [30] S. Zheng et al., "Steve-eye: Equipping LLM-based embodied agents with visual perception in open worlds," in *Proc. 12th Int. Conf. Learn. Representations*, May 2024. [Online]. Available: <https://openreview.net/forum?id=NltzxpG0nZ>
- [31] S. Lifshitz, K. Paster, H. Chan, J. Ba, and S. McIlraith, "STEVE-1: A generative model for text-to-behavior in Minecraft," in *Proc. Int. Conf. Neural Inf. Process. Syst.*, 2024, pp. 69900–69929.
- [32] R. Krishna, D. Lee, L. Fei-Fei, and M. S. Bernstein, "Socially situated artificial intelligence enables learning from human interaction," *Proc. Nat. Acad. Sci.*, vol. 119, no. 39, 2022, Art. no. e2115730119.
- [33] S. Hong et al., "MetaGPT: Meta programming for multi-agent collaborative framework," in *Proc. 12th Int. Conf. Learn. Representations*, 2023.
- [34] Q. Wu et al., "AutoGen: Enabling next-gen LLM applications via multi-agent conversation," in *Proc. Int. Conf. Learn. Representations Workshop Large Lang. Model Agents*, 2024.
- [35] H. Yang, S. Yue, and Y. He, "Auto-GPT for online decision making: Benchmarks and additional opinions," 2023, *arXiv:2306.02224*.
- [36] G. Li, H. A. A. K. Hammoud, H. Itani, D. Khizbulin, and B. Ghanem, "CAMEL: Communicative agents for "mind" exploration of large scale language model society," in *Proc. 37th Int. Conf. Neural Inf. Process. Syst.*, 2023, pp. 51991–52008.
- [37] C.-Y. Lin, "ROUGE: A package for automatic evaluation of summaries," in *Text Summarization Branches Out*. Barcelona, Spain: Assoc. Comput. Linguistics, 2004, pp. 74–81.
- [38] K. Papineni, S. Roukos, T. Ward, and W.-J. Zhu, "BLEU: A method for automatic evaluation of machine translation," in *Proc. 40th Annu. Meeting Assoc. Comput. Linguistics*, 2002, pp. 311–318.
- [39] L. Zheng et al., "Judging LLM-as-a-judge with MT-bench and ChatBot arena," in *Proc. Int. Conf. Neural Inf. Process. Syst.*, 2024, pp. 46595–46623.
- [40] W. Yuan, R. Y. Pang, K. Cho, S. Sukhbaatar, J. Xu, and J. Weston, "Self-rewarding language models," 2024, *arXiv:2401.10020*.
- [41] Y. Shao, L. Li, J. Dai, and X. Qiu, "Character-LLM: A trainable agent for role-playing," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2023, pp. 13153–13187.
- [42] X. Yu et al., "Neeko: Leveraging dynamic LoRA for efficient multi-character role-playing agent," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2024, pp. 12540–12557.
- [43] C. An et al., "L-Eval: Instituting standardized evaluation for long context language models," in *Proc. 62nd Annu. Meeting Assoc. Comput. Linguistics (Volume 1: Long Papers)*, 2024, pp. 14388–14411.
- [44] Meta, "Introducing Meta Llama 3: The most capable openly available LLM to date," 2024. [Online]. Available: <https://ai.meta.com/blog/meta-llama-3/>
- [45] M. I. Ciolacu and P. Svasta, "Education 4.0: AI empowers smart blended learning process with biofeedback," in *Proc. IEEE Glob. Eng. Educ. Conf.*, 2021, pp. 1443–1448.
- [46] Z. Yang, J. Yang, K. Rice, J.-L. Hung, and X. Du, "Using convolutional neural network to recognize learning images for early warning of at-risk students," *IEEE Trans. Learn. Technol.*, vol. 13, no. 3, pp. 617–630, Third Quarter 2020.
- [47] J. C. Tham and G. Verhulsdonck, "Smart education in smart cities: Layered implications for networked and ubiquitous learning," *IEEE Trans. Technol. Soc.*, vol. 4, no. 1, pp. 87–95, Mar. 2023.
- [48] H. Wan, Z. Zhong, L. Tang, and X. Gao, "Pedagogical interventions in SPOCs: Learning behavior dashboards and knowledge tracing support exercise recommendation," *IEEE Trans. Learn. Technol.*, vol. 16, no. 3, pp. 431–442, Jun. 2023.
- [49] H. Singh and S. J. Miah, "Smart education literature: A theoretical analysis," *Educ. Inf. Technol.*, vol. 25, no. 4, pp. 3299–3328, 2020.
- [50] J.-L. Hung, B. E. Shelton, J. Yang, and X. Du, "Improving predictive modeling for at-risk student identification: A multistage approach," *IEEE Trans. Learn. Technol.*, vol. 12, no. 2, pp. 148–157, Second Quarter 2019.
- [51] J. Zhu, X. Ma, and C. Huang, "Stable knowledge tracing using causal inference," *IEEE Trans. Learn. Technol.*, vol. 17, pp. 124–134, 2023.
- [52] A.-B. M. Salem and A. Y. Nikitaeva, "Knowledge engineering paradigms for smart education and learning systems," in *Proc. 42nd Int. Convention Inf. Commun. Electron. Microelectronics*, 2019, pp. 1571–1574.
- [53] D. P. Wangoo and S. N. Reddy, "Smart learning environments framework for educational applications in IoT enabled educational ecosystems: A review on AI based GUI tools for IoT wearables," in *Proc. IEEE 17th India Council Int. Conf.*, 2020, pp. 1–8.
- [54] T. Sutjaritham, H. H. Gharakheili, S. S. Kanhere, and V. Sivaraman, "Experiences with IoT and AI in a smart campus for optimizing classroom usage," *IEEE Internet Things J.*, vol. 6, no. 5, pp. 7595–7607, Oct. 2019.



Xueqiao Zhang received the bachelor's degree in electrical information from the Communication University of Zhejiang, Hangzhou, China, in 2023. He is currently working toward the master's degree in software engineering with the School of Software, Zhejiang University, Hangzhou.

His research interests include multimodal role-playing and multiagents.



Chao Zhang received the bachelor's degree in electrical information from the Wuhan Institute of Technology, Wuhan, China, in 2024. He is currently working toward the master's degree in software engineering with the School of Software, Zhejiang University, Hangzhou, China.

His research interests include role-playing and dialogue systems.



Jianwen Sun received the Ph.D. degree in educational technology from the National Engineering Research Center for E-Learning, Central China Normal University, Wuhan, China, in 2011.

From 2014 to 2016, he was the Director of the Department of Education Informatization, NO.1 Middle School Affiliated to Central China Normal University, Wuhan. From 2016 to 2019, he visited the Israel Institute of Technology, Haifa, Israel; the University of Pittsburgh, Pittsburgh, PA, USA; and the University of Dublin, Dublin, Ireland, for academic exchanges and cooperation. In 2018, he joined the National Engineering Laboratory For Educational Big Data, Central China Normal University. Since 2012, he has been working in educational technology with the National Engineering Research Center for E-Learning, Central China Normal University, where he has also been with the Faculty of Artificial Intelligence in Education, Central China Normal University since 2020. His research interests include perception computing, cognitive graph, and knowledge tracking.



Yi Yang (Senior Member, IEEE) received the Ph.D. degree in computer science from Zhejiang University, Hangzhou, China, in 2010.

He was a Professor and Director of the Recognition, Learning, and Reasoning Lab, Australian Artificial Intelligence Institute, University of Technology Sydney, Ultimo, NSW, Australia. He was a Postdoctoral Researcher at the School of Computer Science, Carnegie Mellon University, Pittsburgh, PA, USA. He is currently a Distinguished Professor with Zhejiang University. His current research interests include machine learning and its applications to multimedia content analysis and computer vision, such as multimedia indexing and retrieval, surveillance video analysis, and video semantics understanding.



Jun Xiao received the Ph.D. degree in computer science and technology from the College of Computer Science, Zhejiang University, Hangzhou, China, in 2007.

He is currently a Professor with the College of Computer Science, Zhejiang University. His current research interests include cross-media analysis, computer vision, and machine learning.



Yawei Luo received the Ph.D. degree in computer application technology from the Huazhong University of Science and Technology, Wuhan, China, in 2020.

From 2020 to 2023, he was a Postdoctoral Researcher at the Collaborative Innovation Center of Artificial Intelligence, College of Computer Science and Technology, Zhejiang University, Hangzhou, China. From 2017 to 2019, he was a visiting Ph.D. Student at the Recognition, Learning, and Reasoning Lab, Australian Artificial Intelligence Institute, University of Technology Sydney, Ultimo, NSW, Australia. He is currently a ZJU 100 young Professor with the School of Software Technology, Zhejiang University. His research interests include knowledge engineering, domain adaptation, and 3-D reconstruction.